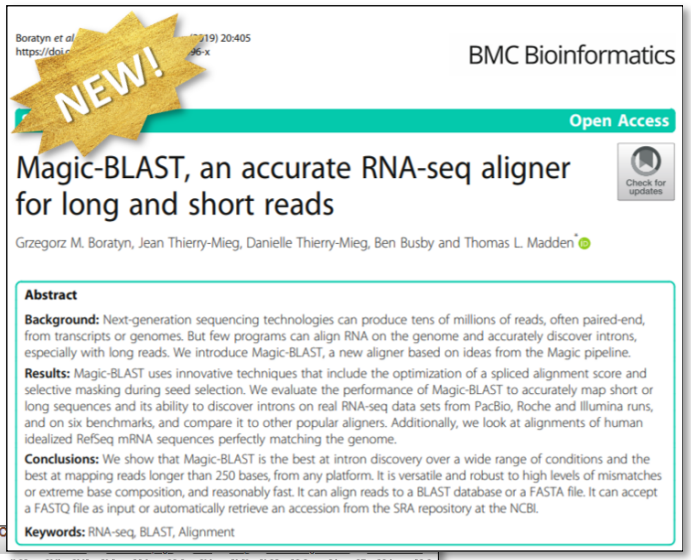
NCBI Resources for NextGen Sequence Analysis

Aligning NextGen data with magic-blast and examining the result in Genome Data Viewer
https://go.usa.gov/xVtK7 | https://www.ncbi.nlm.nih.gov/genome/gdv/
National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Magic-BLAST

A standalone commandline variant of NCBI's famous BLAST sequence search algorithm, this tool was designed for alignment of DNAseq and RNA seq data to a genomic assembly and the results can be stored in several formats.

- **Magic-BLAST Homepage:** http://bit.ly/2M6Bi01
- **Magic-BLAST Publication:** https://go.usa.gov/xVtsM



Genome Data Viewer

The NCBI Genome Data Viewer (GDV) is an interactive genome browser for the exploration and analysis of many types of sequence data in context with curated annotations on an eukaryotic RefSeq genome assembly. This tool facilitates the examination of data from NCBI resources, track hubs, or your own data sets which can be incorporated by uploading a file or streaming your data from a remote location with a URL.

- **GDV Homepage:** https://go.usa.gov/xVtsj
- **GDV Documentation:** https://go.usa.gov/xVtsD

Human Reference Genome Sequences

Magic-BLAST aligns nextgen data to a reference genomic assembly. The Human Reference Genomic Assembly produced by the Genome Reference Consortium (GRC) is a standard for the genomic research community and this assembly is annotated by the NCBI Reference Sequences (RefSeq) project to facilitate understanding and comparison of the data.

- **GRC Homepage for Human:** https://go.usa.gov/xVts9
- **NCBI RefSeq Homepage:** https://go.usa.gov/xVtsn
- **Human Genome Resources Page:** https://go.usa.gov/xVtsP

NextGen Sequence Reads

A source of nextgen data for analysis with Magic-BLAST and visualization with GDV is the NCBI Sequence Read Archive (SRA) database. SRA sequence reads can be downloaded from the NCBI website and are also now available in the AWS and GCP clouds.

- **SRA Homepage:** https://go.usa.gov/xVtHg
- **SRA Search & Download Documentation:** https://go.usa.gov/xVtH4
- **SRA in the Cloud Documentation:** https://go.usa.gov/xVtHD

An Example of Using NCBI's NextGen Sequence Analysis Tools

Finding NextGen Sequence Data at NCBI

The SRA database provides access to submitted nextgen sequence datasets. These can be found by searching SRA directly or from a PubMed record which describes the dataset in a publication.

- **PubMed Record:** https://go.usa.gov/xVtAd

The PubMed SRA link in the "Related Information" section provides a quick way of finding the relevant SRA data records. To download the SRA sequence data, you can use either the SRA Run browser, SRAToolKit or access the data in the AWS or GCP Clouds.

- **SRA Record:** https://go.usa.gov/xVtAs
- **SRA Download Tips:** https://go.usa.gov/xVtA6

Aligning NextGen Sequence Data with Magic-BLAST

The Magic-BLAST GitHub site has guidance on how to download to your own computer, preparation of a reference assembly for alignment, and use of the application with SRA and your own data. Please note that to visualize Magic-BLAST alignments in GDV, you will need to convert the SAM to BAM format which can be done with several open source tools, including SAMtools.

- **Download Magic-BLAST:** http://bit.ly/2IHQn69
- **Making a BLAST Database:** http://bit.ly/2M5dxFQ
- **SRA as Input to Magic-BLAST:** http://bit.ly/329ZkNp

Examining NextGen Sequence data with the Genome Data Viewer

The GDV tool provides both written tutorials and a YouTube video tutorial playlist to assist users in learning to utilize the features and functionality of this versatile visualization tool.

- **GDV Written Tutorials:** https://go.usa.gov/xVtHs
- **GDV YouTube Playlist:** http://bit.ly/35tikIE

GDV provides portlets to "Search" the genomic assembly to zoom-in to a region of interest. "The User Data and Track Hubs" provides a simple way to "Add Remote Files" such as Magic-BLAST alignments. Then, full functionality of GDV enables comparison with curated and annotated data by configuring the data "Tracks" shown.

Example Magic-BLAST alignments:

- https://go.usa.gov/xVtHz
- https://go.usa.gov/xVtHh